МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение  
высшего образования «Ижевский государственный технический  
университет имени М. Т. Калашникова»

Институт «Информатика и вычислительная техника»

Кафедра «Защита информации в компьютеризированных системах»

**ОТЧЕТ**

по лабораторной работе №3

по дисциплине «Разработка и эксплуатация защищенных автоматизированных систем»

на тему «Применение регулярных выражений для решения задач информационной безопасности»

Выполнили:  
студент гр. С17-361-1 Буров К.А.

Проверил:  
ст. преп. П. А. Кудашев

Ижевск 2021

# **Цель работы**

Изучить прицнип работы регулярных выражений и применить из для парсинга, нормализации и классификации логов, поиска нежелательного контента в каком-либо источнике и выявления попыток подделки доменных имен.

# **Задачи**

1. Найти подходящие для парсинга логи.
2. Разработать регулярные выражения для поиска информации в логах
3. Распарсить и нормализовать логи.
4. Классифицировать логи.
5. Разработать регулярные выражения для поиска нежелательного контента.
6. Найти нежелательный контент в источнике.
7. Разработать механизм обнаружения попыток подделки доменных имен.
8. Протестировать разработанный механизм

# **Ход работы**

1. Найти подходящие для парсинга логи.

Для парсинга были выбраны логи домашнего роутера, которые были экспортированы в текстовый файл.

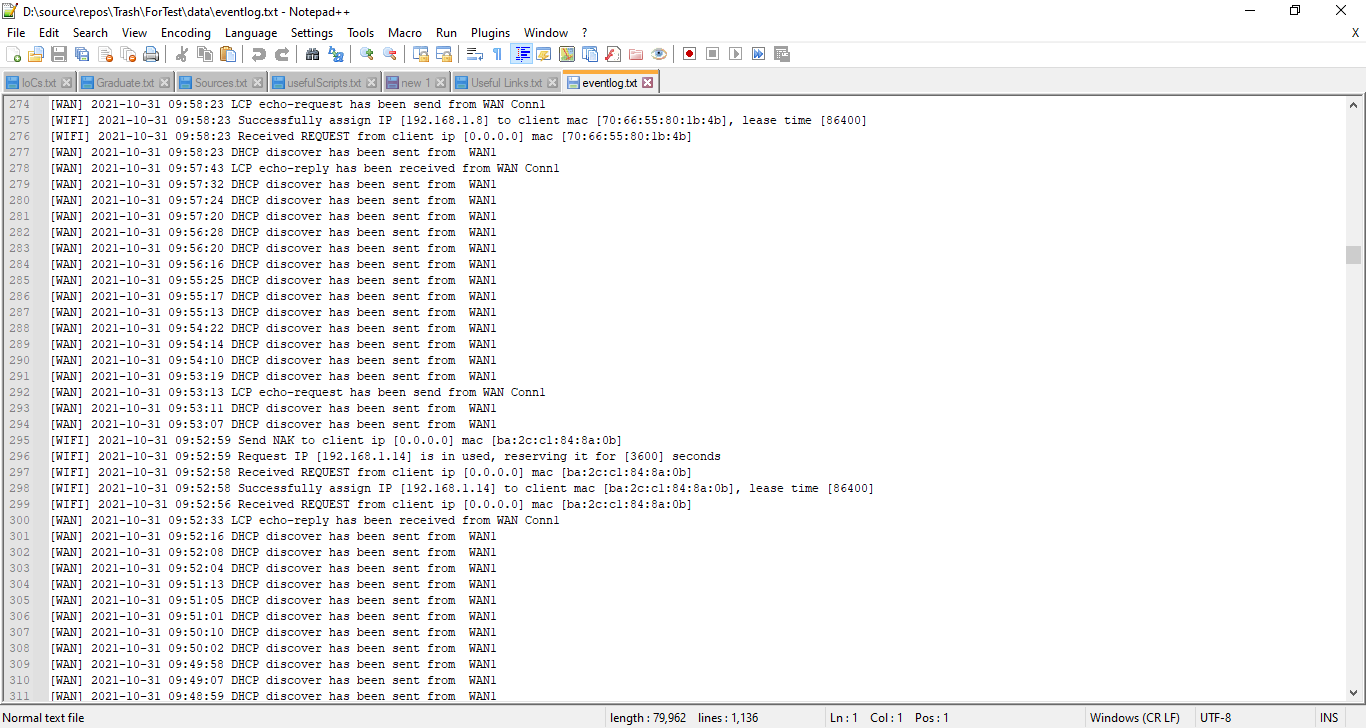


Рисунок - Логи домашнего роутера

1. Разработать регулярные выражения для поиска информации в логах

Для поиска инфомрации в логах роутера были разработаны регулярные выражения для поиска:

* Даты и времени
* Категории события
* Протокола
* IP-адреса
* MAC-адреса

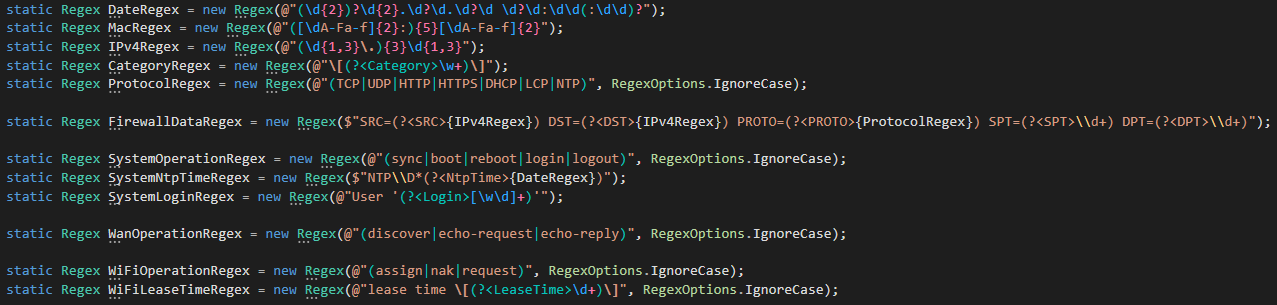


Рисунок - Регулярные выражения для парсинга логов

Отдельное внимание стоит уделить парсингу даты и времени так как форматов их записи достаточно много. Исходя из этого были разработаны наборы регулярных выражений для даты и для времени и генератор, который путем попарной конкатенации выстаривает множество всех возможных вариантов даты и времени.

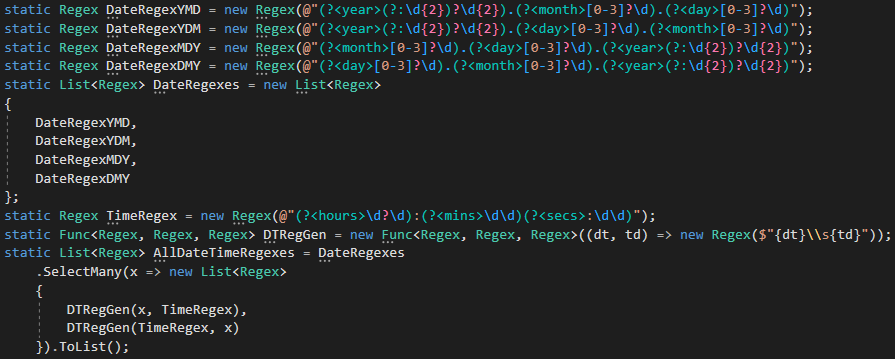


Рисунок - Регулярные выражения для парсинга даты и времени

1. Распарсить и нормализовать логи.
2. Классифицировать логи.

События из логов были разобраны с помощью регулярок в массив объектов, классифицированы по категории события и сериализованы в JSON

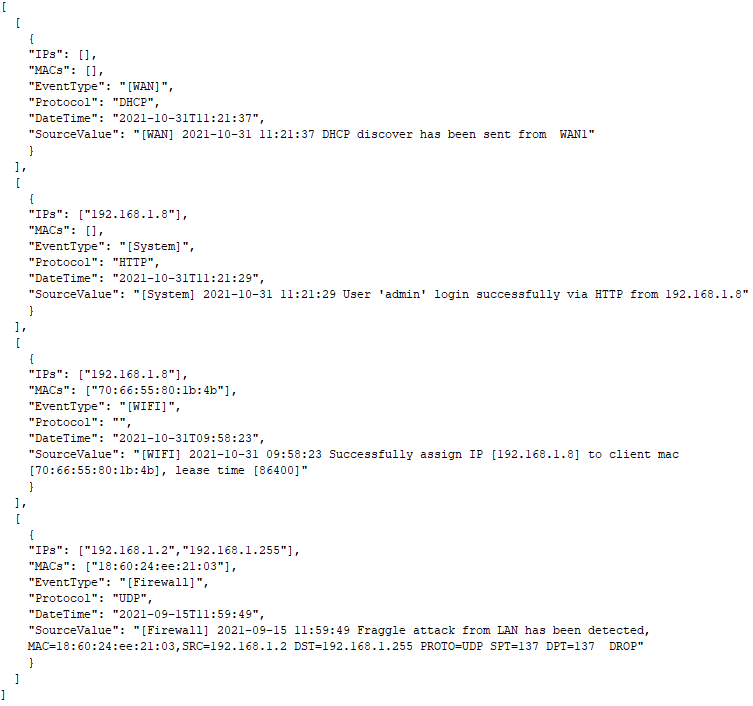


Рисунок – Логи роутера после парсинга, нормализации и классификации

1. После классификации можно сделать более точный парсинг по категориям событий. Для этого воспользуемся регулярными выражениями, специфичными для каждой из категорий событий. После более подробного парсинга событий получаем следующий результат:

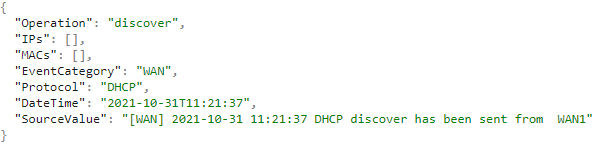


Рисунок 5 - Разобранные логи категории WAN

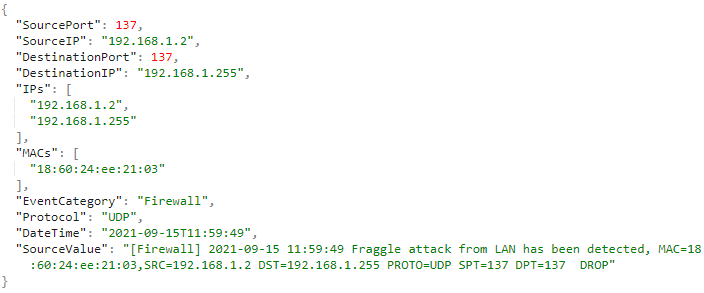


Рисунок 6 - Разобранные логи категории Firewall

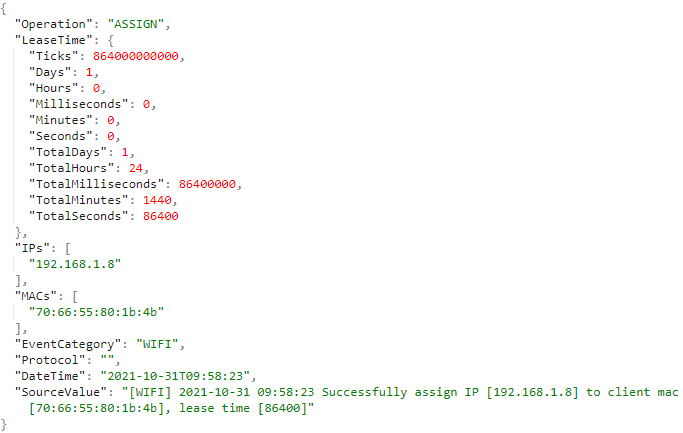


Рисунок 7 - Разобранные логи категории WIFI

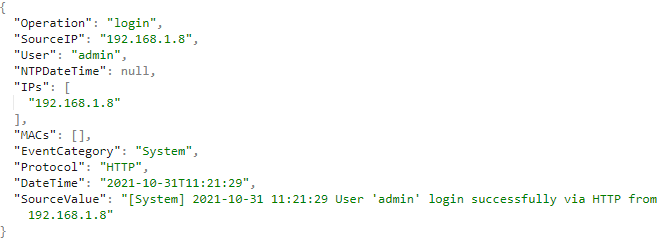


Рисунок 8 - Разобранные логи категории System

1. Разработать регулярные выражения для поиска нежелательного контента.

Регулярные выражения для поиска нежелательного контента формируются динамически из списка вызывающих подозрение слов по следующей логике: если слово короче 5 символов, то регулярное выражение проверяет эквивалентность этому слову, в противном случае из слова удаляются последние два символа и в конец добавляется регулярное выражение для вариативности окончаний (0-4 словообразующих символа).

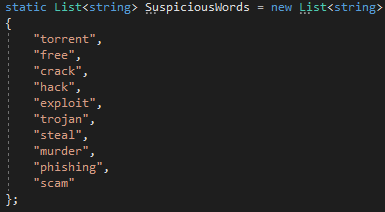


Рисунок – Пример подозрительных слов для поиска

1. Найти нежелательный контент в источнике.

Поиск нежелательного контента будет производится на перехваченном HTTP трафике

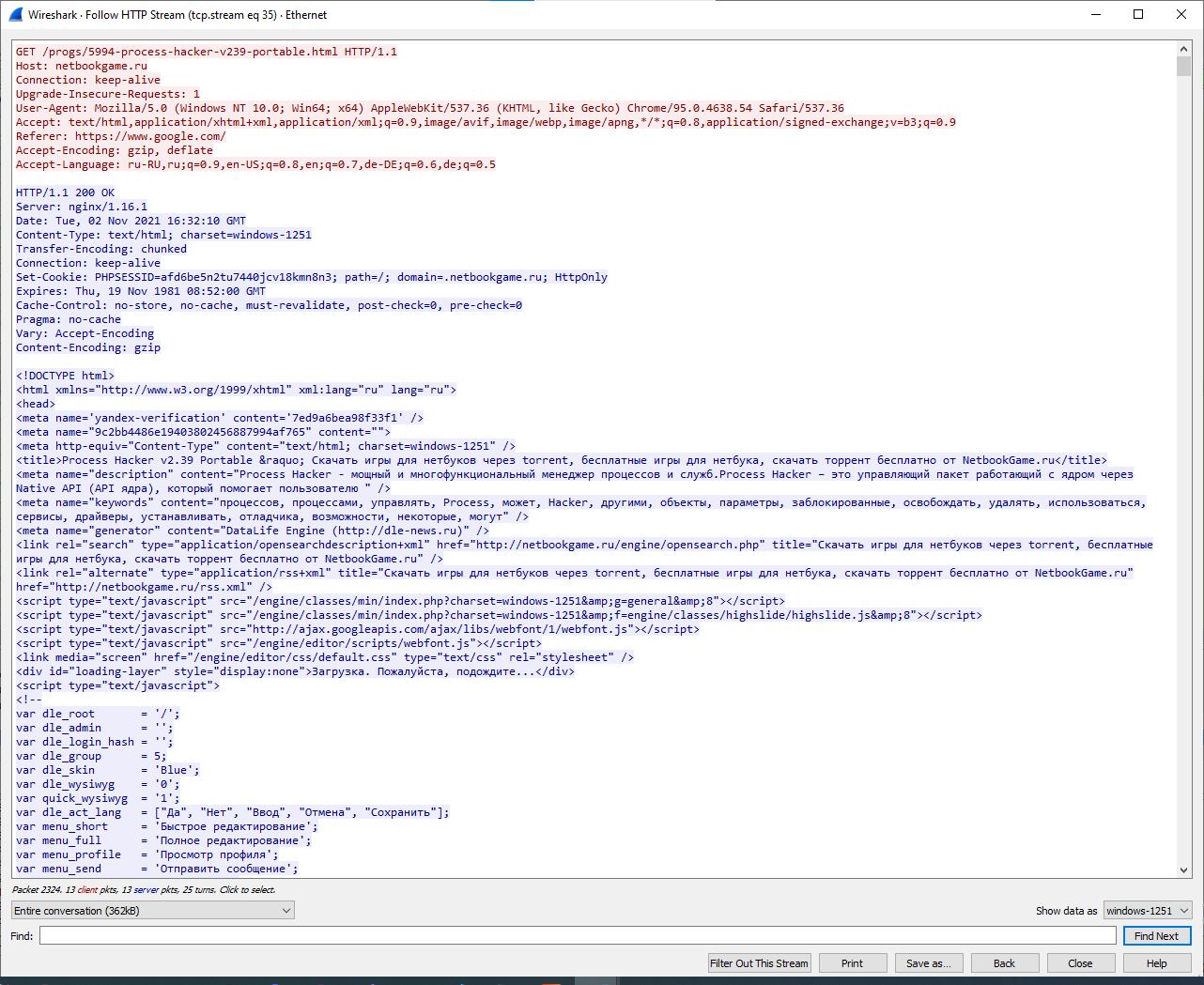


Рисунок – Часть перехваченного трафика

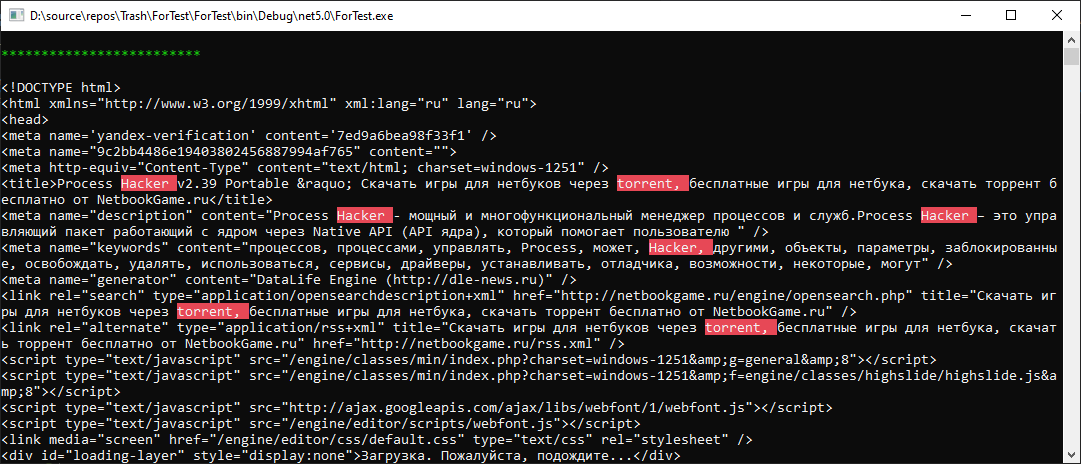


Рисунок - Пример поиска подозрительного контента

1. Разработать механизм обнаружения попыток подделки доменных имен.

Перед началом разработки самого механизма проверки был составлен словарь, сопоставляющий обычные латинские символы с символами, которые похожи на них и могут использованы для маскировки зловредного домена:

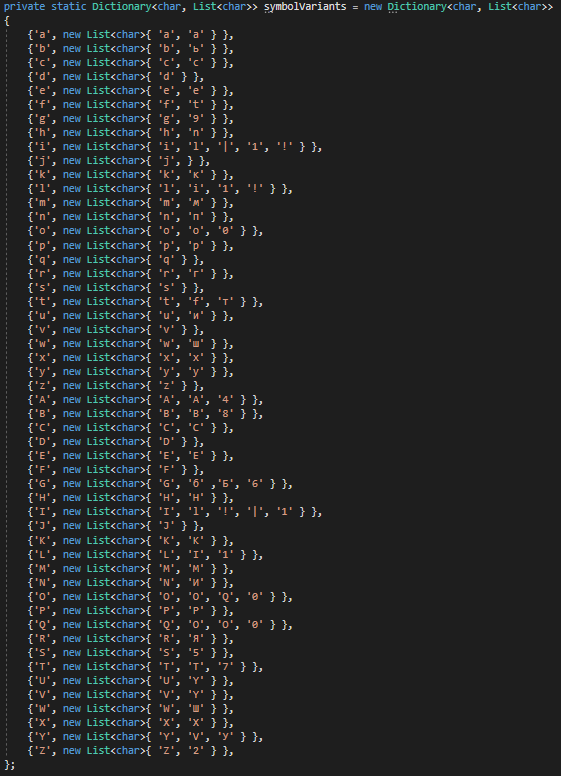


Рисунок - Словарь подмены символов

Также был составлен тестовый список популярных доверенных доменов:



Рисунок - Список доверенных доменов

Механизм валидации на основании перечня доверенных доменов и словаря подмены символов составляет регулярные выражения, путем замены каждого конкретного символа на перечень символов способных его заменить и также учитывает то, что символы могут дублироваться. По полученным регулярным выражениям производится проверка на соответствие проверяемого домена маске, а затем проверка на эквивалентность доверенному домену. В случае, когда домен подходит по маске, но не эквивалентен доверенному домену он считается вредоносным.

1. Протестировать разработанный механизм

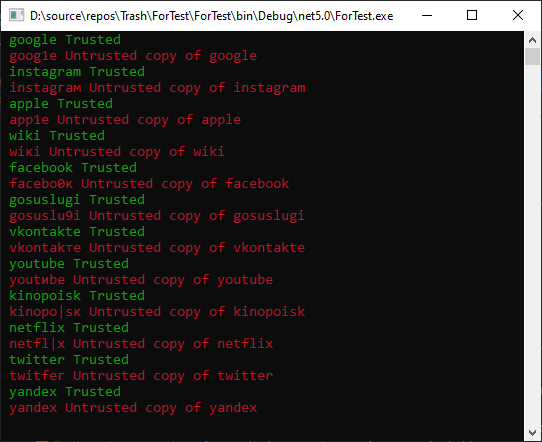


Рисунок - Тестирование механизма проверки домена

**Вывод:** в рамках лабораторной работы был изучен принцип работы регулярных выражений. Регулярные выражения были применены для парсинга, нормализации и классификации логов роутера. Также была проведена проверка контента на предмет нежелательных слов. Был разработан и протестирован механизм проверки подделки доменных имен основанный на регулярных выражениях. На основании проведенной работы можно сделать вывод, что регулярные выражения это мощный инструмент для анализа текстовой информации, который позволяет из большого массива разнородных данных извлечь ценную информацию.